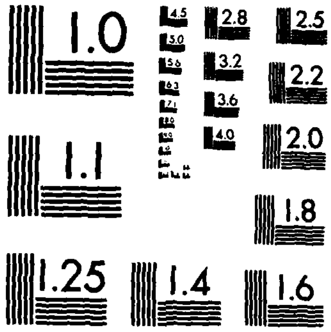
AD-A164 094 A BERRY-ESSEEN THEOREM FOR WEIGHTED U-STATISTICS(U) 1/1
JOHNS HOPKINS UNIV BALTIMORE MD DEPT OF MATHEMATICAL
SCIENCES J C WIERMAN JAN 83 TR-365 N00014-79-C-0801

UNCLASSIFIED F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

7

A Berry-Esséen Theorem for Weighted U-Statistics*

(Short Title: Weighted U-Statistics)

Technical Report No. 365

by

John C. Wierman
Department of Mathematical Sciences
The Johns Hopkins University

Revised, January, 1983

Contract N00014-79-C-0801

ABSTRACT

A Berry-Esseen theorem is proved for weighted U-statistics, assuming certain growth conditions are satisfied by sums of the weights. The result is proved using the Fourier-analytic techniques of Chan & Wierman (1977) and Callaert & Janssen (1978). Keywords: random variables, distribution functions; asymptotic normality, symmetric functions.

AMS Subject Classification: 60F07
Key words and phrases: Berry-Esseen bounds, weighted U-statistic, reduced U-statistic.

*Research sponsored in part by the National Science Foundation under grant MCS-8118229 and by the Office of Naval Research under contract N00014-79-C-0801.

1. Introduction

Let $X_1, X_2, \ldots, X_n$, $n \geq 2$ be i.i.d. random variables with common distribution function F. Let h be a symmetric function of r variables such that $h(X_1, \ldots, X_r)$ has mean zero and such that $E[h(X_1, \ldots, X_r) | X_1] = g(X_1)$ has a positive variance. Hoeffding [11] introduced the U-statistic

$$H_n = \binom{n}{r}^{-1} \sum_{\underline{i} \in C} h(X_{i_1}, \ldots, X_{i_r}),$$

where $\sum_{\underline{i} \in C}$ denotes summation over the set C of combinations $\underline{i} = i_1, \ldots, i_r$ of integers in $\{1,2,\ldots,n\}$. Hoeffding proved the asymptotic normality of U-statistics. An investigation of the rate of convergence to normality begun by Grams and Serfling [9] and continued by Bickel [1] and Chan & Wierman [4], resulted in the Berry Esseen theorem for U-statistics by Callaert and Janssen [3]. They obtained the rate of convergence $O(n^{-\frac{1}{2}})$ assuming a finite absolute third moment for the kernel $h(X_1, \ldots, X_r)$. Recently, Helmers and Van Zwet [10], for the case of r=2, have relaxed the assumption to $E|g(X_1)|^3 < \infty$ and $E|h(X_1, X_2)|^t < \infty$ for some $t > 5/3$.

For a symmetric function $w(i_1, \ldots, i_r)$ on $(I_n)^r$, where $I_n = \{1,2,\ldots,n\}$, satisfying the condition that $w(i_1, \ldots, i_r) = 0$ if $i_j = i_k$ for some $j \neq k$, we define the weighted U-statistic

$$U_n = \sum_{\underline{i} \in C} w(i_1, \ldots, i_r) h(X_{i_1}, \ldots, X_{i_r}).$$

Little is known concerning the asymptotic properties of such statistics, as noted by Serfling [15]. For kernels of degree r=2, Brown and Kildea [2] considered statistics of the form $S_n = \sum_{(i,j) \in C_{K,n}} h(X_i, X_j)$, where k is fixed and for each n, $C_{K,n}$ is a collection of pairs (i,j) with $1 \leq i < j \leq n$ balanced in such a manner that each positive integer less than or equal to n

is present in exactly 2K pairs in $C_{K,n}$. These statistics are called balanced incomplete U-statistics or reduced U-statistics, and are clearly a special case of the weighted U-statistic with weights of 0 or 1 only. Brown and Kildea show that S_n, properly standardized, is asymptotically normal. Estimates based on reduced U-statistics are asymptotically equivalent to those based on the corresponding U-statistics, but require far fewer steps to compute. Brown and Kildea also obtain asymptotic normality in some cases when the balancing condition is relaxed.

Sievers [17] considered the simple linear regression model $Y_i = \alpha + \beta x_i + e_i$, $1 \leq i \leq n$, where α and β are unknown parameters, $x_1,\ldots,x_n$ are known regression scores, and $e_1,\ldots,e_n$ are i.i.d. random variables. He considered inferences for β based on a weighted rank statistic defined by

$$T_\beta = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ij}\phi(Y_i - \alpha - \beta x_i, Y_j - \alpha - \beta x_j)$$

where $\phi(u,v) = 1$ if $u \leq v$ and 0 if $u > v$. The weights are arbitrary, except that $a_{ij} = 0$ if $x_i = x_j$. Note that the when the slope parameter has value β, then T_β is a weighted U-statistic. Sievers proved asymptotic normality of T_β under restrictions on the weights a_{ij}, and developed tests and confidence intervals for the value of the slope parameter β based on T_β.

Shapiro and Hubert [16] consider weighted U-statistics with kernels of order r=2, and proved asymptotic normality if $E[h(X_1,X_2)^2] < \infty$ and

$$\sum_{i\neq j} w_{ij}^2 \Big/ \sum_{k=1}^{n} w_{k\cdot n}^2 \rightarrow 0$$

and

$$\max_{1\leq i\leq n} w_{i\cdot n}^2 \Big/ \sum_{k=1}^{n} w_{k\cdot n}^2 \rightarrow 0$$

where $w_{i\cdot n} = \sum_{j=1}^{n} w_{ij}$. This result is then used to obtain asymptotic normality of permutation statistics of interest in biometry (Mantel and Valand [14]), geography (Cliff and Ord [5]) and clustering studies (Hubert and Schultz [12].

Kepner and Robinson [13] considered weighted sums of multivariate functions with kernel of order k, and generalized the asymptotic normality results of Brown and Kildea [2] and Shapiro and Hubert. Note that the results of :hese papers and the present paper are valid when the kernel h and weight f :tion w are replaced by sequences h_n and w_n satisfying the conditions assu

Let $U_n = \sum_{\underline{i} \in C} w(i_1,\ldots,i_r)h(X_{i_1},\ldots,X_{i_r})$

where $\sum_{\underline{i} \in C}$ denotes the sum of over all combinations $\underline{i} = \{i_1,\ldots,i_r\}$ of integers from $\{1,2,\ldots,n\}$. Introduce the function g by $g(X_{i_1}) = E[h(X_{i_1},\ldots,X_{i_r})|X_{i_1}]$, and the sums of weights

$$w_i = \sum_{\underline{i} \ni i} w(\underline{i}) = \sum_{\substack{i_2 \\ \{i,i_2,\ldots i_r\} \in C}} \cdots \sum_{i_r} w(i,i_2,\ldots,i_r)$$

and $$w_{ij} = \sum_{\underline{i} \ni i,j} w(\underline{i}) = \sum_{\substack{i_3 \\ \{i,j,i_3,\ldots,i_r\} \in C}} \cdots \sum_{i_r} w(i,j,i_3,\ldots,i_r).$$

Let $r_n = \sum_{i=1}^{n} w_i^3$, $s_n^2 = \sum_{i=1}^{n} w_i^2 = \sigma_g^{-2}\hat{\sigma}_n^2$, and $t_n = \sum_{i=1}^{n}\sum_{j=i+1}^{n} w_{ij}^2$.

The projection of U_n is given by

$$\hat{U}_n = \sum_{i=1}^{n} E[U_n|X_i]$$

$$= \sum_{i=1}^{n} \sum_{\substack{\underline{i} \ni i \\ \underline{i} \in C}} w(i_1,\ldots,i_r)E[h(X_{i_1},\ldots,X_{i_r})|X_i]$$

$$= \sum_{i=1}^{n} \left\{ \sum_{\substack{\underline{i} \ni i \\ \underline{i} \in C}} w(i_1,\ldots,i_r) g(X_i) \right\}$$

$$= \sum_{i=1}^{n} w_i g(X_i)$$

or alternatively

$$\hat{U}_n = \sum_{i=1}^{n} E[U_n | X_i] = \sum_{\underline{i} \in C} \left\{ w(i_1,\ldots,i_r) \sum_{i=1}^{n} E[h(X_{i_1},\ldots,X_{i_r}) | X_i] \right\}$$

$$= \sum_{\underline{i} \in C} w(i_1,\ldots,i_r) [g(X_{i_1}) + \ldots + g(X_{i_r})]$$

Let $\sigma_g^2 = \mathrm{Var}\ (g(X_i))$, $\sigma_h^2 = \mathrm{Var}\ (h(X_1,\ldots,X_r))$, and $\sigma_n^2 = \mathrm{Var}(U_n)$. Calculate $\hat{\sigma}_n^2 = \mathrm{Var}(\hat{U}_n) = \sum_{i=1}^{n} w_i^2\ \mathrm{Var}(g(X_i)) = \sigma_g^2 \sum_{i=1}^{n} w_i^2$.

Three conditions on the weights are required for the statement of the result.

Condition (1): There exists $B < 1$ for which

$$\max_{1 \le i \le n} w_1^2 \le (Bs_n^2/r)\Delta(s_n^6/r_n^2) \quad \text{for all } n \ge r + 1$$

Condition (2): $$\frac{\max w_{ij}^2}{\max w_1^2} \le \frac{1}{3} n^{-1} r_n^{11/3} s_n^{-9} [t_n \log(\sigma_h \sigma_g^{-1} s_n^5 r_n^{-2} t_n^{1/2})]^{-1}$$

Condition (3): $t_n \le Cr_n/s_n$ for some C, $0 < C < \infty$.

<u>Theorem</u>: If $h(X_1,\ldots,X_r)$ has finite absolute third moment and the weights satisfy Conditions (1), (2), and (3), then

$$\sup_{x \in \mathbb{R}} | P(\sigma_n^{-1} U_n \le x) - \Phi(x) | = O(r_n/s_n^3), \text{ as } n \to \infty.$$

The most restrictive condition on the weights is Condition (2), which is derived from the characteristic function bounds in the Fourier analytic approach to the Berry-Esseen result. With Conditions (1) and (2) satisfied, the theorem holds for $U_n/\hat{\sigma}_n$. Condition (3) permits replacement of $\hat{\sigma}_n$ by σ_n.

The present paper generalizes the result of Callaert and Janssen [3], since if $w(\underline{i}) = 1$ for all i, conditions (1), (2), and (3) are satisfied, and in this case the rate of convergence is $O(r_n/s_n^3) = O(n^{-\frac{1}{2}})$. For the case of unequal weights satisfying $0 < A \leq w(\underline{i}) \leq B$ for all $\underline{i}$, the Theorem applies and provides an $O(n^{-\frac{1}{2}})$ rate of convergence. In fact, one sufficient condition for the convergence rate $O(n^{-\frac{1}{2}})$ is

$$(*)\qquad \frac{\max s_{ij}}{\min w_{ij}} \leq B,$$

which holds for the above-mentioned cases in this paragraph. One may observe from Conditions (1), (2) and (3) that the bound on the convergence rate depends on the weights only through their sums w_{ij}, so individual weights may differ greatly without violating the hypotheses of the Theorem.

2. Proof of Theorem

Denote $(U_n - \hat{U}_n)/\hat{\sigma}_n$ by Δ_n. Note that

$$\Delta_n = \hat{\sigma}_n^{-1} \sum_{\underline{i} \in C} w(i_1,\ldots,i_r) Y_{i_1,\ldots,i_r}$$

where $Y_{i_1,\ldots,i_r} = h(X_{i_1},\ldots,X_{i_r}) - g(X_{i_1}) - \ldots - g(X_{i_r})$.

Split Δ_n into two parts Δ_n' and $\Delta_n'' = \Delta_n - \Delta_n'$, with

$$\Delta_n' = \sum_{i_1=1}^{c_n} \sum_{i_2=i_1+1}^{c_n} \cdots \sum_{i_r=i_r+1}^{c_n} w(i_1,\ldots,i_r) Y_{i_1}) Y_{i_1,\ldots,i_r}.$$

Restrictions on the choice of c_n are found which provide the rate of convergence $O(r_n/s_n^3)$ for bounds on several terms to be estimated. Condition (2) insures the existence of a choice of c_n which satisfies all of these restrictions simultaneously. [This corresponds to the analysis of order bounds for c_n and d_n in section 3 of Callaert and Janssen [3].]

For any sequence a_n of constants, an elementary calculation gives

$$\sup_x |P(U_n/\hat{\sigma}_n \le x) - \Phi(x)|$$

$$\le \sup_x |P(S_n + \Delta_n' \le x) - \Phi(x)| + P(|\Delta_n''| \ge a_n) + O(a_n)$$

Then, letting ϕ_X denote the characteristic function of random variable X, for x>0,

$$\int_0^{\varepsilon s_n^3/r_n} t^{-1} |e^{-t^2/2} - \phi_{S_n+\Delta_n'}(t)| dt$$

$$\le \int_0^{\varepsilon s_n^3/r_n} t^{-1} |e^{-t^2/2} - \phi_{S_n}(t)| dt + \int_0^{\varepsilon s_n^3/r_n} t^{-1} |\phi_{S_n}(t) - \phi_{S_n+\Delta_n'}(t)| dt.$$

Since S_n is a sum of independent random variables with finite absolute third moments, a standard Berry-Esseen argument (see e.g. Feller [8], p.544) yields

$$\int_0^{\varepsilon s_n^3/r_n} t^{-1} |e^{-t^2/2} - \phi_{S_n}(t)| dt \le C_1 \nu_3 \sigma_g^{-3} \, r_n/s_n^3$$

for an absolute constant C_1, where $\nu_3 = E|h(X_1,\ldots,X_r)|^3$, and we may take $\varepsilon = \sigma_g^3/\nu_3$.

The majority of the proof determines the bound for the remaining integral. Writing η for the characteristic function of $g_n(X_1)$, with ε as above, we have

$$|\eta(\theta)| \le e^{-\frac{1}{3}\theta^2\sigma_g^2} \qquad \text{for } |\theta| \le \varepsilon\sigma_g^{-1}.$$

Begin by estimating

$$|\phi_{S_n}(t) - \phi_{S_n+\Delta_n'}(t)|$$

$$= |E[e^{itS_n}(1-e^{it\Delta_n'})]|$$

$$\le |E[e^{itS_n} \, it\Delta_n']| + \tfrac{1}{2} t^2 E(\Delta_n')^2,$$

and note that by independence,

$$|E[e^{itS_n}\Delta_n']|$$

$$= \frac{1}{\hat{\sigma}_n}\left|\sum_{\substack{\underline{i}\in C \\ i,j\le c_n \forall j}} w(i_1,\ldots,i_r)E[e^{it\hat{\sigma}_n^{-1}\sum_{k\ne i_j \forall j} w_k g(X_k)}]\right.$$

$$\left.\times E\left[e^{it\hat{\sigma}_n^{-1}(w_{i_1}g(X_{i_1})+\ldots+w_{i_r}g(X_{i_r}))}Y_{i_1\ldots i_r}\right]\right| .$$

For a fixed combination $\underline{i} = \{i_1,\ldots,i_r\}$, assuming Condition (1) holds, for $0 < t < \varepsilon\, s_n^3/r_n$,

$$\left|E\left[e^{it\hat{\sigma}_n^{-1}\sum_{k=i_j\, j} w_k g(X_k)}\right]\right| = \prod_{k\ne i_1,\ldots,i_n}\left|\eta\left(w_k\hat{\sigma}_n^{-1}t\right)\right|$$

$$\le e^{-\frac{t^2}{3}\hat{\sigma}_n^{-2}\left(\sum_{k\ne i_1,\ldots,i_r} w_k^2\right)\sigma_g^2}$$

$$= e^{-\frac{t^2}{3}(1-B)}$$

since $0 < t < \varepsilon s_n^3/r_n < \varepsilon s_n/\max_{1\le i\le n} w_i$ implies $w_k\,\hat{\sigma}_n^{-1}t \le \varepsilon\sigma_g^{-1}$ for all k.

Also, for each fixed $\underline{i}$, since $E[f(X_{i_j})Y_{i_1},\ldots,i_r] = 0$ for any bound Borel measurable function f,

$$\left|E\left[e^{it\hat{\sigma}_n^{-1}\sum_{j=1}^{r} w_{i_j}g(X_{i_j})}Y_{i_1,\ldots,i_r}\right]\right|$$

$$= \left|E\left[\left\{e^{it\hat{\sigma}_n^{-1}\sum_{j=1}^{r} w_{i_j}g(X_{i_j})} - 1 - it\hat{\sigma}_n^{-1}\sum_{j=1}^{n} w_{i_j}g(X_{i_j})\right\}Y_{i_1,\ldots,i_r}\right]\right| ,$$

$$\leq t\hat{\sigma}_n^{-2} \sum_{i_j} \sum_{i_k} w_{i_j} w_{i_k} E|g(X_{i_j}) g(X_{i_k}) Y_{i_1,\ldots,i_r}|$$

$$\leq (r+1)\nu_3 t^2 \hat{\sigma}_n^{-2} \left(\sum_{j=1} w_{i_j} \right)^2 .$$

Combine these estimates to obtain

$$|Ee^{itS_n} \Delta_n'|$$

$$\leq (r+1)\nu^3 \hat{\sigma}_n^{-3} t^2 e^{-\frac{t^2}{3}(1-B)} \sum_{\substack{i_j \leq c_n \forall j \\ \underline{i} \in C}} \{w(i_1,\ldots,i_r)(w_{i_1}+\ldots+w_{i_r})^2\} .$$

To bound this sum, write

$$\sum_{\underline{i} \in C} w(i_1,\ldots,i_r)[w_{i_1}^2+\ldots+w_{i_r}^2 + 2w_{i_1}w_{i_2}+\ldots+2w_{i_{r-1}}w_{i_r}].$$

First,

$$\sum_{\underline{i} \in C} w(i_1,\ldots,i_r) w_{i_1}^2$$

$$= \sum_{i_1=1}^{n} \{w_{i_1}^2 \sum_{\underline{i} \ni i_1} w(i_1,\ldots,i_r)\}$$

$$= \sum_{i_1=1}^{n} w_{i_1}^3$$

and similarly for each of the squared term's contributions. For the cross-product terms, by Hölder's inequality,

$$\sum_{\underline{i} \in C} w(i_1,\ldots,i_r) w_{i_1} w_{i_2}$$

$$= \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \{ w_{i_1} w_{i_2} \sum_{\substack{\underline{i} \ni i_1, i_2 \\ \underline{i} \in C}} w(i_1, i_2, \ldots, i_r) \}$$

$$= \sum_{i_1=1}^{n} \sum_{i_2=i_1+1}^{n} w_{i_1} w_{i_2} w_{i_1 i_2}$$

$$\leq \left[\sum_{i_1=1}^{n} \sum_{i_2=i_1+1}^{n} w_{i_1}^2 w_{i_1 i_2} \right]^{1/2} \left[\sum_{i_1=1}^{n} \sum_{i_2=i_1+1}^{n} w_{i_2}^2 w_{i_1 i_2} \right]^{1/2}$$

$$\leq \sum_{i}^{n} w_i^3$$

There are $\binom{r}{2}$ cross-product sums, each with a coefficient of 2, so combining the bounds, the overall sum is bounded by

$$r^2 \sum_{i=1}^{n} w_i^3 .$$

Hence for $d_n \leq \varepsilon \frac{s_n^3}{r_n}$,

$$\int_0^{d_n} t^{-1} \; |E[e^{itS_{\Delta'}}]| \, dt$$

$$\leq \frac{r^2 (r+1) \nu_3}{\hat{\sigma}_n^3} \sum_{i=1}^{n} w_i^3 \int_0^{d_n} t^2 \, e^{\frac{t^2 (1-B)}{3}} \, dt$$

$$\leq \frac{r^2 (r+1) \nu_3}{\sigma_g^3} \frac{r_n}{s_n^3} \frac{3}{4} \sqrt{\pi} \, (1-B)^{-3/2}$$

independent of the choice of d_n. Note also that the choice of c_n played no role in the computation of this bound. To bound $E[(\Delta'_n)^2]$, note that

$$E[Y_{i_1,\ldots,i_r} Y_{j_1,\ldots,j_r}] = 0$$

if the combinations $\underline{i}$ and $\underline{j}$ contain zero or one common indices. [See Grams and Serfling [9]. Otherwise,

$$E[Y_{i_1,\ldots,i_r} Y_{j_1,\ldots,j_r}] \leq (r+1)^2 \sigma_h^2$$

Then $\hat{\sigma}_n^2 \; E[(\Delta'_n)^2]$

$$= E\left[\sum_{\underline{i}\in C} \sum_{\underline{j}\in C} w(\underline{i})w(\underline{j})Y_{\underline{i}}Y_{\underline{j}}\right]$$

$$= \sum_{\substack{\underline{i},\underline{j} \\ |\underline{i}\cap \underline{j}|\geq 2}} w(\underline{i})w(\underline{j})\; E[Y_{\underline{i}}Y_{\underline{j}}]$$

$$\leq (r+1)^2\sigma_h^2 \sum_{i_1=1}^{n} \sum_{i_2=i_1+1}^{n} \left(\sum_{\underline{i}\supseteq\{i_1,i_2\}} \sum_{\underline{j}\supseteq\{j_1,i_2\}} w_{\underline{i}} w_{\underline{j}}\right)$$

$$= (r+1)^2\sigma_h^2 \sum_{i_1=1} \sum_{i_2=i_1+1} w_{i_1,i_2}^2$$

$$= (r+1)^2\sigma_h^2 t_n .$$

The choice of d_n id determined by the bound for $E(\Delta')^2$. Choosing $d_n = r_n^{\frac{1}{2}} t_n^{-\frac{1}{2}} s_n^{-\frac{1}{2}}$, the bound becomes

$$\frac{1}{2}\, E(\Delta')^2 \int_0^{d_n} t\; dt$$

$$\leq \tfrac{1}{4}(r+1)^2\sigma_h^2\hat{\sigma}_n^{-2}\; t_n\, d_n^2$$

$$\leq \tfrac{1}{4}(r+1)^2\sigma_h^2\sigma_g^{-2}\; r_n/s_n^3$$

The estimates above provide the bound required for $|P(\hat{\sigma}_n^{-1}U_n \le x) - \Phi(X)|$ for all n such that $\varepsilon s_n^3/r_n \le d_n$.

For the case when $d_n < r_n/s_n^3$, write

$$\left|E\left[e^{itS}(1-e^{it\Delta'})\right]\right|$$

$$= \left|E\left[e^{it\hat{\sigma}_n^{-1}\sum_{k>c_n} w_k g(X_k)}\right]\right| \left|E\left[e^{it\hat{\sigma}_n^{-1}\sum_{k\le c_n} w_k g(X_k)}(1-e^{it\Delta'})\right]\right|$$

$$\le \left|\prod_{k>c_n} \eta(\hat{\sigma}_n^{-1} w_k t)\right| E\left[\left|1-e^{it\Delta'}\right|\right]$$

$$\le \left(e^{-\frac{1}{3}\hat{\sigma}_n^{-2}\sum_{k>c_n} w_k^2 t^2 \zeta_1}\right) E\left[\left|1-e^{it\Delta'}\right|\right]$$

$$\le t\,E|\Delta'|\; e^{-\frac{1}{3}\sigma_n^{-2}\sum_{k>c_n} w_k^2 t^2 \sigma_g^2}$$

The bound for $E|\Delta'|$ is obtained from Lyapunov's inequality ([6], p. 47) and the previous bound for $E(\Delta')^2$:

$$E|\Delta'| \le E^{\frac{1}{2}}(\Delta')^2 \le (r+1)\sigma_h \sigma_n^{-1} t_n^{\frac{1}{2}}.$$

Choose c_n so that

$$(*) \quad \sum_{k>c_n} w_k^2 \ge 3\hat{\sigma}_n^{-2}\sigma_g^{-1} d_n^{-2} \log(\sigma_h \sigma_g^{-1} s_n^5 r_n^{-2} t_n^{\frac{1}{2}}).$$

Then

$$\int_{d_n}^{\varepsilon s_n^3/r_n} t^{-1} |\phi_S(t) - \phi_{S+\Delta'}(t)|\,dt$$

$$\leq (r+1)\,\sigma_h \hat{\sigma}_n^{-1} t_n^{\frac{1}{2}} \int_{d_n}^{\varepsilon s_n^3/r_n} e^{-\frac{1}{3}\hat{\sigma}_n^2 \zeta_1^{-1} t^2 \sum_{k>c_n} w_k^2}\, dt$$

$$\leq (r+1)\,\sigma_h \hat{\sigma}_n^{-1} t_n^{\frac{1}{2}} \left(\varepsilon \frac{s_n}{r_n}\right) \Big/ \left(e^{-\frac{1}{3}\hat{\sigma}_n^2 \zeta_1^{-1} \sum_{k>c_n} w_k^2} \right)$$

$$= \varepsilon (r+1)\,\sigma_h \hat{\sigma}_g^{-1} r_n / s_n^3 .$$

Note that inequality (*) is satisfied if

$$(n-c_n) \min_{1\leq i\leq n} w_i^2 \geq 3\sigma_g^{-2} s_n \sigma_n^2 r_n^{-1} t_n \log(\sigma_h \sigma_g^{-1} s_n^5 r_n^{-2} t_n^{\frac{1}{2}}),$$

providing a lower bound for $n-c_n$ to be used later in the proof.

To handle Δ'', define ξ_j by

$$\dot{\hat{\sigma}}_n \Delta'' = \sum_{j=c_n+1}^{n} \{ \sum_{\underline{i} \ni j} w_{\underline{i}} Y_{\underline{i}} \} = \sum_{j=c_n+1}^{n} \xi_j .$$

Since $E[\xi_{j+1} | \xi_i,\ i \leq j] = 0$ a.s. for all j, the ξ_j are martingale summands, and by optional skipping, $V_k = \sum_{j=c_n+1}^{c_n+k} \xi_j$ forms a martingale, $k = 1,2,\ldots,n-c_n$. By a theorem of Dharmadhikari, Fabian, and Jogdeo [7], for $k = n-c_n$,

$$E|V_{n-c_n}|^3 \leq 2^{12} (n-c_n)^{3/2} \max_{c_n+1\leq j\leq n} E|\xi_j|^3 .$$

However, for fixed $j \geq c_n+1$,

$$W_k = \sum_{i=1}^{k} \{ \sum_{\underline{i}\ni j,\, i} w_{\underline{i}} Y_{\underline{i}} \} , \quad k=1,2,\ldots,j-1$$

is also a martingale. A second application of the theorem of Dharmadhikari, Fabian, and Jogdeo [7] yields

$$E|\xi_j|^3 = E|W_{j-1}|^3 \le 2^{12}(j-1)^{3/2} \max_{1\le i\le j-1} E|\sum_{\underline{i}\ni j,i} w_{\underline{i}} Y_{\underline{i}}|^3.$$

Now

$$E|\sum_{\underline{i}\ni j,i} w_{\underline{i}} Y_{\underline{i}}|^3 \le \sum\sum\sum_{\underline{i}_1,\underline{i}_2,\underline{i}_3\ni j,i} w_{\underline{i}_1} w_{\underline{i}_2} w_{\underline{i}_3} E\, |Y_{\underline{i}_1} Y_{\underline{i}_2} Y_{\underline{i}_3}|$$

$$\le (r+1)^3 \nu_3 \sum\sum\sum_{\underline{i}_1,\underline{i}_2,\underline{i}_3 \ni j,i} w_{\underline{i}_1} w_{\underline{i}_2} w_{\underline{i}_3}$$

$$= (r+1)^3 \nu_3\, w_{ij}^3.$$

Therefore,

$$E|\Delta''|^3 \le 2^{24}(r+1)^3 \nu_3 (n-c_n)^{3/2} n^{3/2} [\max_{i,j} w_{ij}^3] \hat{\sigma}_n^3.$$

By the Markov inequality,

$$P(|\Delta''| \ge a_n) \le a_n^{-3} E|\Delta''|^3.$$

Taking $a_n = \left[(n-c_n)^{3/2} n^{3/2} \hat{\sigma}_n^{-3} \max w_{ij}^3\right]^{1/4}$ yields

$$P(|\Delta''| \ge a_n) \le 2^{24}(r+1)^3 \nu_3 a_n.$$

If c_n is chosen so that

$$n-c_n \le \frac{r_n^{8/3}}{n\hat{\sigma}_n^6} \; \frac{1}{\max w_{i,j}^2},$$

then $a_n \le \dfrac{r_n}{s_n^3}$.

Finally, if both conditions concerning $n-c_n$ may be satisfied simultaneously, the $O(r_n/s_n^3)$ rate of convergence is obtained for $U_n/\hat{\sigma}_n$. This provides the condition

$$\frac{\max w_{i,j}^2}{\min w_{\underline{i}}^2} \le \frac{1}{3n} \frac{r_n^{11/3}}{s_n^9}\, t_n \log(\sigma_h \sigma_g^{-1} s_n^5 r_n^{-2} t_n^{1/2}).$$

Note that the condition depends on the weights only through their sums $w_{\underline{i}}$ and $w_{\underline{i}\underline{j}}$.

To replace $\hat{\sigma}_n$ by σ_n, note that

$$\mathrm{Var}(U_n) = \sum_{\underline{i}} \sum_{\underline{j}} w_{\underline{i}} w_{\underline{j}}\, E[h(X_{i_1},\ldots,X_{i_r})h(X_{j_1},\ldots,X_{j_r})]$$

$$= \sum_{\substack{\underline{i}\ \underline{j} \\ |\underline{i}\cap\underline{j}|=1}} w_{\underline{i}} w_{\underline{j}} \sigma_g^2 + \sum_{\substack{\underline{i}\ \underline{j} \\ |\underline{i}\cap\underline{j}|\geq 2}} w_{\underline{i}} w_{\underline{j}}\, E[X_{i_1},\ldots,X_{i_r})h(X_{j_1},\ldots,X_{j_r})] .$$

$$= \sum_{i=1}^{n} w_i^2 \sigma_g^2 + \sum_{\substack{\underline{i}\ \underline{j} \\ |\underline{i}\cap\underline{j}| \geq 2}} w_{\underline{i}} w_{\underline{j}}\, E(h(X_{i_1},\ldots,X_{i_r})h(X_{j_1},\ldots,X_{j_r})]$$

implies $|\sigma_n^2 - \hat{\sigma}_n^2| \leq \sigma_h^2 \sum\limits_{\substack{\underline{i}\ \underline{j} \\ |\underline{i}\cap j| \geq 2}} w_{\underline{i}} w_{\underline{j}} = \sigma_h^2 \sum\limits_i \sum\limits_j w_{ij}^2$.

Therefore

$$\left|\frac{\sigma_n}{\hat{\sigma}_n} - 1\right| \leq \left|\frac{\sigma_n}{\hat{\sigma}_n} - 1\right| \left|\frac{\sigma_n}{\hat{\sigma}_n} + 1\right|$$

$$= \left|\frac{\sigma_n^2 - \hat{\sigma}_n^2}{\hat{\sigma}_n^2}\right|$$

$$\leq \frac{\sigma_h^2 \sum\limits_i \sum\limits_i w_{ij}^2}{\sigma_g^2 \sum\limits_i w_i^2} .$$

If $t_n \leq c\, r_n/s_n$, then there exists a constant K such that

$$P(\hat{\sigma}_n^{-1}U_n \le (1-Kr_n/s_n^3)x) \le P(\sigma_n^{-1}U_n \le x) \le P(\hat{\sigma}_n^{-1}U_n \le (1+Kr_n/s_n^3)x)$$

for all positive real numbers x, with a similar inequality for $x < 0$. By the assertion of the theorem,

$$\begin{aligned}
&P(\sigma_n^{-1}U_n \le x) \\
&\le P(\hat{\sigma}_n^{-1}U_n \le (1+Kr_n/s_n^3)x) \\
&\le \Phi((1+Kr_n/s_n^3)x) + Lr_n/s_n^3 \\
&\le \Phi(x) + Lr_n/s_n^3 + (2\pi)^{-\frac{1}{2}}e^{-x^2/2}Kr_n/s_n^3 \\
&\le \Phi(x) + Mr_n/s_n^3
\end{aligned}$$

Using similar reasoning for the lower bound, the replacement of $\hat{\sigma}_n$ by σ_n is shown to preserve the convergence rate r_n/s_n^3.

3. References

[1] Bickel, P.J. (1974). Edgeworth expansions in nonparametric statistics. Ann. Statist. 2, 1-20.

[2] Brown, B.M. and Kildea, D.G. (1978). Reduced U-statistics and the Hodges-Lehmann estimator. Ann. Statist. 6, 828-835.

[3] Callaert, H. and Janssen, P. (1978). The Berry-Esseen Theorem for U-statistics. Ann. Statist. 6, 417-421.

[4] Chan, Y.K. and Wierman, J.C. (1977). On the Berry-Esseen Theorem for U-statistics. Ann. Probab. 5, 136-139.

[5] Cliff, A.D. and Ord, J.K. (1973) Spatial Autocorrelation. Pion, London.

[6] Chung, K.L. (1974) A Course in Probability Theory. Academic Press, New York.

[7] Dharmadhikari, S.W., Fabian, V. and Jogdeo, K. (1968). (1968). Bounds on the moments of martingales. Ann. Math. Statist. 39, 1719-1723.

[8] Feller, W. (1966). An Introduction to Probability Theory and Its Applications. 2. Wiley, New York.

[9] Grams, W.F. and Serfling, R.J. (1973). Convergence rates for U-statistics. Ann. Statist. 1, 153-160.

[10] Helmers, R. and Van Zwet, W.R. (1981) The Berry-Esséen bound for U-statistics. Statistical theory and related topics, S.S. Gupta, editor.

[11] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. Ann. Math. Statist. 19, 293-325.

[12] Hubert, L. and Schultz, J. (1976) Quadratic assignment as a general data analysis strategy. British J. Math. Statist. Psychology 29, 190-241.

[13] Kepner, J.L. and Robinson, D.H. (1982). On the asymptotic normality of weighted sums of multivariate functions. University of Florida Technical Report 158.

[14] Mantel, N. and Valand, R.S. (1970). A technique of nonparametric multivariate analysis. *Biometrics* 27, 547-558.

[15] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

[16] Shapiro, C.P. and Hubert, L. (1979). Asymptotic normality of permutation statistics derived from weighted sums of bivariate functions. *Ann. Statist.* *7*, 783-794.

[17] Sievers, G.L. (1978) Weighted rank statistics for simple linear regression. *JASA* *73*, 628-631.

END

FILMED

3-86

DTIC